

Patent Number :

CN1264741 A 20000830 DW2000-64 C12N-015/55 *

AP: 2000CN-0111778 20000302

Priority Details :

2000CN-0111778 20000302

IPC s :

C12N-015/55 C07K-016/40 C12N-009/16 C12N-015/63 C12Q-001/68

Abstract :

CN1264741 A

NOVELTY - The present invention discloses a new human adenosine-diphosphate ribosylglucose hydrolase (hARGHase) protein expressed in human liver cancer cells and its coding sequence. The process of preparing the protein and nucleic acid sequence and the process for detecting the nucleic acid sequence and polypeptide of the human hARGHase in specimen are also disclosed. (Dwg.0/0)

Manual Codes :

CPI: B04-E03E B04-N02 B11-C08E D05-C03C D05-C07 D05-H12A D05-H18

Update Basic :

2000-64

[19]中华人民共和国国家知识产权局

[51]Int. Cl⁷

C12N 15/55

[12] 发明专利申请公开说明书

C12N 9/16 C12N 15/63

C07K 16/40 C12Q 1/68

[21] 申请号 00111778.5

[43]公开日 2000年8月30日

[11]公开号 CN 1264741A

[22]申请日 2000.3.2 [21]申请号 00111778.5

[71]申请人 国家人类基因组南方研究中心

地址 201203 上海市浦东张江高科技园区郭守敬路351号

[72]发明人 李能干 钱斌治 高新

肖华胜 陈竺 韩泽广

[74]专利代理机构 复旦大学专利事务所

代理人 陆飞

权利要求书1页 说明书19页 附图页数0页

[54]发明名称 一种新的人二磷酸腺苷核糖基葡萄糖水解酶蛋白及其编码序列

[57]摘要

本发明提供了一种在人体肝癌组织中表达的新的hARGHase蛋白及其编码序列,本发明还提供了该蛋白和核酸序列的制备方法,以及在样品中检测人hARGHase核酸序列和多肽的方法。

ISSN 1000-8427

知识产权出版社出版

权利要求书

1. 一种分离出的 DNA 分子, 其特征在于, 它包括: 编码具有人 hARGHase 蛋白质活性的多肽的核苷酸序列, 而且所述的核苷酸序列与 SEQ ID NO. 6 中从核苷酸第 51-1142 位的核苷酸序列有至少 70% 的同源性; 或者所述的核苷酸序列能在中度严谨条件下与 SEQ ID NO. 6 中从核苷酸第 51-1142 位的核苷酸序列杂交。
2. 如权利要求 1 所述的 DNA 分子, 其特征在于, 所述的序列编码具有 SEQ ID NO. 7 所示的氨基酸序列的多肽。
3. 如权利要求 1 所述的 DNA 分子, 其特征在于, 该序列具有 SEQ ID NO. 6 中从核苷酸第 51-1142 位的核苷酸序列。
4. 一种分离出的人 hARGHase 蛋白多肽, 其特征在于, 它包括: 具有 SEQ ID NO. 7 氨基酸序列的多肽、或其保守性变异多肽、或其活性片段, 或其活性衍生物。
5. 如权利要求 4 所述的多肽, 其特征在于, 该多肽是具有 SEQ ID NO. 7 序列的多肽。
6. 一种载体, 其特征在于, 它包含权利要求 1 所述的 DNA。
7. 一种用权利要求 6 所述载体转化的宿主细胞, 其特征在于它包括原核细胞和真核细胞。
8. 一种产生具有人 hARGHase 蛋白质活性的多肽的方法, 特征在于其步骤如下:
 - (1) 将编码具有人 hARGHase 蛋白质活性的多肽的核苷酸序列可操作地连于表达调控序列, 形成人 hARGHase 蛋白表达载体, 所述的核苷酸序列与 SEQ ID NO. 6 中从核苷酸第 51-1142 位的核苷酸序列有至少 70% 的同源性;
 - (2) 将步骤(1)中的表达载体转入宿主细胞, 形成人 hARGHase 蛋白的重组细胞;
 - (3) 在适合表达人 hARGHase 蛋白多肽的条件下, 培养步骤(2)中的重组细胞;
 - (4) 分离出具有人 hARGHase 蛋白质活性的多肽。
9. 一种能与权利要求 7 所述的人 hARGHase 蛋白多肽特异性结合的抗体, 其特征在于它包括多克隆抗体和单克隆抗体。
10. 一种核酸分子, 其特征在于, 它包含权利要求 1 所述的 DNA 分子中 8-100 个连续核苷酸。
11. 一种用于检测样品中是否存在人 hARGHase 核苷酸序列的方法, 其特征在于它包括用权利要求 10 所述探针与样品进行杂交, 然后检测探针是否发生了结合, 该样品是 PCR 扩增后的产物, 其中 PCR 扩增引物对应于人 hARGHase 核苷酸编码序列, 并可位于该编码序列的两侧或中间, 引物长度为 15~50 个核苷酸。

说明书

一种新的人二磷酸腺苷核糖基葡萄糖水解酶蛋白及其编码序列

本发明涉及分子生物学、酶学、生理学以及基因工程等领域。具体地，本发明涉及一种在人类肝癌组织中表达的 hARGHase 蛋白及其核酸序列。本发明还涉及该蛋白和核酸序列的制备方法和用途。

二磷酸腺苷核糖基葡萄糖水解酶 (ADP-ribosylglycohydrolase, ARGHase) 参与生物的新陈代谢，主要功能是水解 ADP 偶联的核糖基葡萄糖。(Nature 392, 353-358 (1998))。

我们在人肝癌组织中克隆到此酶的人类同源基因 hARGHase，同源性研究表明 hARGHase 有与此酶相同或者相近的功能。

在本发明被公布之前，尚未有任何公开或报道过人类 hARGHase 蛋白序列以及编码这个蛋白的核酸序列。

本发明的第一目的就是提供一种新的人基因 hARGHase (Genbank Accession No. AF212236)，该基因是一个 hARGHase 蛋白基因。

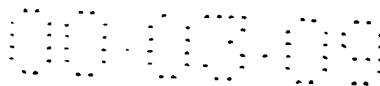
本发明的第二目的是提供一种新的人蛋白 hARGHase。

本发明的第三目的是提供一种利用重组技术生产上述的新的人 hARGHase 蛋白和核酸序列的方法。

本发明还提供了这种人 hARGHase 蛋白多肽和编码序列的应用。

在本发明的一个方面，提供了一种分离出的 DNA 分子，该分子包括：编码具有人 hARGHase 蛋白质活性的多肽的核苷酸序列，所述的核苷酸序列与 SEQ ID NO. 6 中从核苷酸第 51-1142 位 DNA 分子的核苷酸序列有至少 70% 的同源性；或者所述的核苷酸序列能在中度严谨条件下与 SEQ ID NO. 6 中从核苷酸第 51-1142 位的核苷酸序列杂交。较佳地，所述的序列编码具有 SEQ ID NO. 7 所示的氨基酸序列的多肽。更佳地，所述的序列具有 SEQ ID NO. 6 中从核苷酸第 51-1142 位的核苷酸序列。

在本发明的另一方面，提供了一种分离出的人 hARGHase 蛋白质多肽，它包括：具有 SEQ ID NO. 7 氨基酸序列的多肽、或其保守性变异多肽、或其活性片段，或其活性衍生物。较佳地，该多肽是具有 SEQ ID NO. 7 序列的多肽。



在本发明的另一方面，还提供了一种载体，它包含上述的 DNA 分子。

在本发明的另一方面，还提供了一种用上述载体转化的宿主细胞。在一个实例中该宿主细胞是大肠杆菌；在另一实例中，该宿主细胞是真核细胞。

在本发明的另一方面，还提供了一种产生具有人 hARGHase 蛋白质活性的多肽的方法，其步骤如下：

(1) 将编码具有人 hARGHase 蛋白活性的多肽的核苷酸序列可操作地连于表达调控序列，形成人 hARGHase 蛋白表达载体，所述的核苷酸序列与 SEQ ID NO. 6 中从核苷酸第 51-1142 位的核苷酸序列有至少 70% 的同源性；

(2) 将步骤(1)中的表达载体转入宿主细胞，形成人 hARGHase 蛋白的重组细胞；

(3) 在适合表达人 hARGHase 蛋白多肽的条件下，培养步骤(2)中的重组细胞；

(4) 分离出具有人 hARGHase 蛋白活性的多肽。

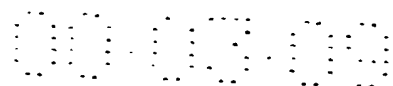
较佳地，在该方法中使用的核酸序列具有 SEQ ID NO. 6 中第 51-1142 位的序列。

本发明还提供了与 hARGHase 蛋白多肽特异性结合的抗体。

在本发明中，“分离的”、“纯化的”或“基本纯的”DNA 是指，该 DNA 或片段已从天然状态下位于其两侧的序列中分离出来，还指该 DNA 或片段已经与天然状态下伴随核酸的组份分开，而且已经与在细胞中伴随其的蛋白质分开。

在本发明中，术语“人 hARGHase 蛋白(或多肽)编码序列”指编码具有人 hARGHase 蛋白活性的多肽的核苷酸序列，如 SEQ ID NO. 6 中第 51-1142 位核苷酸序列及其简并序列。该简并序列是指，位于 SEQ ID NO. 6 序列的编码框第 51-1142 位核苷酸中，有一个或多个密码子被编码相同氨基酸的简并密码子所取代后而产生的序列。由于密码子的简并性，所以与 SEQ ID NO. 6 中第 51-1142 位核苷酸序列同源性低至约 70% 的简并序列也能编码出 SEQ ID NO. 7 所述的序列。该术语还包括能在中度严紧条件下，更佳的在高度严紧条件下与 SEQ ID NO. 6 中从核苷酸第 51-1142 位的核苷酸序列杂交的核苷酸序列。该术语还包括与 SEQ ID NO. 6 中从核苷酸第 51-1142 位的核苷酸序列的同源性至少 70%，较佳地至少 80%，更佳地至少 90%，最佳地至少 95% 的核苷酸序列。

该术语还包括能编码具有与天然的人 hARGHase 相同功能的蛋白的、SEQ ID NO. 6 中开放阅读框序列的变异形式。这些变异形式包括(但不限于)：若干个(通常为 1-90 个，较佳地 1-60 个，更佳地 1-20 个，最佳地 1-10 个)核苷酸的缺失、插入和/或取代，以及在 5' 和/或 3' 端添加数个(通常为 60 个以内，较佳地为 30 个以内，更佳地为 10 个以内，最佳地为 5 个以内)核苷酸。



在本发明中，“基本纯的”蛋白质或多肽是指其至少占样品总物质的至少 20%，较佳地至少 50%，更佳地至少 80%，最佳地至少 90%（按干重或湿重计）。纯度可以用任何合适的方法进行测量，如用柱层析、PAGE 或 HPLC 法测量多肽的纯度。基本纯的多肽基本上不含天然状态下的伴随其的组分。

在本发明中，术语“人 hARGHase 蛋白或多肽”指具有人 hARGHase 蛋白活性的 SEQ ID NO. 7 序列的多肽。该术语还包括具有与天然人 hARGHase 相同功能的、SEQ ID NO. 7 序列的变异形式。这些变异形式包括（但并不限于）：若干个（通常为 1-50 个，较佳地 1-30 个，更佳地 1-20 个，最佳地 1-10 个）氨基酸的缺失、插入和/或取代，以及在 C 末端和/或 N 末端添加一个或数个（通常为 20 个以内，较佳地为 10 个以内，更佳地为 5 个以内）氨基酸。例如，在本领域中，用性能相近或相似的氨基酸进行取代时，通常不会改变蛋白质的功能。又比如，在 C 末端和/或 N 末端添加一个或数个氨基酸通常也不会改变蛋白质的功能。该术语还包括人 hARGHase 蛋白的活性片段和活性衍生物。

本发明的人 hARGHase 多肽的变异形式包括：同源序列、保守性变异体、等位变异体、天然突变体、诱导突变体、在高或低的严紧条件下能与入 hARGHase DNA 杂交的 DNA 所编码的蛋白、以及利用抗人 hARGHase 多肽的抗血清获得的多肽或蛋白。本发明还提供了其他多肽，如包含人 hARGHase 多肽或其片段的融合蛋白。除了几乎全长的多肽外，本发明还包括人 hARGHase 多肽的可溶性片段。通常，该片段具有人 hARGHase 多肽序列的至少约 10 个连续氨基酸，通常至少约 30 个连续氨基酸，较佳地至少约 50 个连续氨基酸，更佳地至少约 80 个连续氨基酸，最佳地至少约 100 个连续氨基酸。

在本发明中，“人 hARGHase 保守性变异多肽”指与 SEQ ID NO. 7 的氨基酸序列相比，有至多 10 个，较佳地至多 8 个，更佳地至多 5 个氨基酸被性质相似或相近的氨基酸所替换而形成多肽。这些保守性变异多肽最好根据表 1 进行替换而产生。

表 1

最初的残基	代表性的取代	优选的取代
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Lys; Arg	Gln
Asp (D)	Glu	Glu

Cys (C)	Ser	Ser
Gln (Q)	Asn	Asn
Glu (E)	Asp	Asp
Gly (G)	Pro; Ala	Ala
His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe	Leu
Leu (L)	Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Leu; Val; Ile; Ala; Tyr	Leu
Pro (P)	Ala	Ala
Ser (S)	Thr	Thr
Thr (T)	Ser	Ser
Trp (W)	Tyr; Phe	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala	Leu

发明还包括人 hARGHase 蛋白或多肽的类似物。这些类似物与天然人 hARGHase 多肽的差别可以是氨基酸序列上的差异，也可以是不影响序列的修饰形式上的差异，或者兼而有之。这些多肽包括天然或诱导的遗传变异体。诱导变异体可以通过各种技术得到，如通过辐射或暴露于诱变剂而产生随机诱变，还可通过定点诱变法或其技术得到。类似物还包括具有不同于天然 L-氨基酸的残基(如 D-氨基酸)的类似物，以及具有非天然存在的或合成的氨基酸(如 β 、 γ -氨基酸)的类似物。应理解，本发明的多肽并不限于上述例举的代表性的多肽。

修饰(通常不改变一级结构)形式包括：体内或体外的多肽的化学衍生形式如乙酰化或羧基化。修饰还包括糖基化，如那些在多肽的合成和加工中或进一步加工步骤中进行糖基化修饰而产生的多肽。这种修饰可以通过将多肽暴露于进行糖基化的酶(如哺乳动物的糖基化酶或去糖基化酶)而完成。修饰形式还包括具有磷酸化氨基酸残基(如磷酸酪氨酸，磷酸丝氨酸，磷酸苏氨酸)的序列。还包括被修饰从而提高其抗蛋白水解性能或优化了溶解性能的多肽。

在本发明中，可选用本领域已知的各种载体，如市售的载体，包括质粒，粘粒等。在生产本发明的人 hARGHase 多肽时，可以将人 hARGHase 编码序列可操作地连

于表达调控序列，从而形成人 hARGHase 蛋白表达载体。

如本文所用，“可操作地连于”指这样一种状况，即线性 DNA 序列的某些部分能够影响同一线性 DNA 序列其他部分的活性。例如，如果信号肽 DNA 作为前体表达并参与多肽的分泌，那么信号肽(分泌前导序列)DNA 就是可操作地连于多肽 DNA；如果启动子控制序列的转录，那么它是可操作地连于编码序列；如果核糖体结合位点被置于能使其翻译的位置时，那么它是可操作地连于编码序列。一般，“可操作地连于”意味着相邻，而对于分泌前导序列则意味着在阅读框中相邻。

在本发明中，术语“宿主细胞”包括原核细胞和真核细胞。常用的原核宿主细胞的例子包括大肠杆菌、枯草杆菌等。常用的真核宿主细胞包括酵母细胞、昆虫细胞和哺乳动物细胞。较佳地，该宿主细胞是真核细胞，如 CHO 细胞、COS 细胞等。

本发明还提供了对人 hARGHase 特异性结合的抗体，包括多克隆抗体和单克隆抗体。

在本发明中，可以使用一系列本领域已知的方法来制备针对人 hARGHase 特异的抗体。例如，将提纯的人 hARGHase 基因产物或它的抗原片段注射入动物体内以产生多克隆抗体。同样，表达人 hARGHase 或它的抗原片段的细胞也可以用来对动物致免疫而产生抗体。根据本发明制备的抗体也可以是单克隆抗体，这些单克隆抗体可以用杂交瘤技术制备(例如，Kohler et al., Nature 256: 495, 1975; Kohler et al., Eur. J. Immunol. 6: 511, 1976; Kohler et al., Eur. J. Immunol. 6: 292, 1976)。本发明的抗体包括可以阻抑人 hARGHase 功能的抗体，也可以是不影响人 hARGHase 功能的抗体。每一类抗体都可以通过对人 hARGHase 基因产物的片段或功能域致免疫而产生，而人 hARGHase 基因产物及其片段可以用重组方法产生或用多肽合成仪进行合成。与非修饰形式的人 hARGHase 基因产物结合的抗体，可以通过用在原核细胞例如 E. coli 中产生的基因产物来免疫动物而得到。与翻译后修饰形式如糖基化或磷酸化蛋白或多肽结合的抗体，可以通过用在真核细胞如酵母或昆虫细胞中产生的基因产物的来免疫动物而得到。

本发明的人 hARGHase 抗体可以用来鉴定表达人 hARGHase 蛋白或多肽的细胞，如 Jurkat T 细胞。例如，可以用一种可检测的分子例如荧光素异硫氰酸(FITC)来标记人 hARGHase 特异抗体，然后让人 hARGHase 特异抗体与细胞样品接触，再用荧光显微镜或流式细胞仪检测出与人 hARGHase 特异抗体结合的细胞。

除了在细胞表面检测人 hARGHase 外，还可以用 Western 印迹技术分析该蛋白质。细胞裂解液可以从培养细胞或取自病人的组织标本如肝癌组织中提取，并溶解在含有去污剂的裂解缓冲液中。然后用 SDS 聚丙烯酰胺凝胶电泳分离细胞提取物(同

时将提纯的人 hARGHase 多肽作为阳性对照), 接着通过电泳杂交将其转移到硝酸纤维素上。为了用 Western 印迹免疫探测人 hARGHase 多肽, 可以使用典型的抗体结合检测方法, 例如放射自显影或碱性磷酸酶检测方法。并可使用免疫接种血清或不相关的单克隆抗体作为非特异反应的对照。

还可利用 Northern 印迹法技术分析人 hARGHase 基因产物的表达, 即分析人 hARGHase 的 RNA 转录物在细胞中的存在与否和数量

人 hARGHase DNA 的 Northern 印迹分析和人 hARGHase 特异抗体的 Western 印迹分析可以联合使用, 以证实人 hARGHase 在生物样本中的表达。人 hARGHase DNA 还可以用于 Southern 印迹分析或原位杂交分析, 以将该基因定位于染色体上, 并可进行遗传连锁分析以找出其它可能的疾病相关基因。

此外, 本发明还提供了一种可用作探针的核酸分子, 该分子通常具有人 hARGHase 核苷酸编码序列的 8-100 个连续核苷酸, 较佳地具有 15-50 个连续核苷酸。该探针可用于检测样品中是否存在编码人 hARGHase 的核酸分子。

本发明还提供了检测样品中是否存在人 hARGHase 核苷酸序列的方法, 它包括用上述的探针与样品进行杂交, 然后检测探针是否发生了结合。较佳地, 该样品是 PCR 扩增后的产物, 其中 PCR 扩增引物对应于人 hARGHase 核苷酸编码序列, 并可位于该编码序列的两侧或中间。引物长度一般为 15-50 个核苷酸。

此外, 根据本发明的人 hARGHase 核苷酸序列和氨基酸序列, 可以在核酸同源性或表达蛋白质的同源性基础上, 筛选人 hARGHase 同源基因或同源蛋白。

为了得到与人 hARGHase 基因相关的人 cDNAs 或基因组 DNAs 的点阵, 可以用 DNA 探针筛选人 cDNA 或基因组 DNA 文库, 这些探针是在低严谨条件下, 用 ^{32}P 对人 hARGHase 的全部或部分做放射活性标记而得的。最适合于筛选的 cDNA 文库是来自人肝癌组织的文库。来自参与内分泌的其它人体组织或特定人体细胞株的 cDNA 文库也可用于筛选目的。构建来自感兴趣的细胞或者组织的 cDNA 文库的方法是分子生物学领域众所周知的。另外, 许多这样的 cDNA 文库也可以购买到, 例如购自 Clontech, Palo Alto, Cal.。这种筛选方法可以识别与人 hARGHase 相关的基因家族的核苷酸序列。

根据核苷酸相似性筛选人 hARGHase 同源物可以按如下方法完成。人体肝癌组织 cDNA 文库, 例如 Clontech Cat. #74XX (Clontech, Palo Alto, Cal.) 可以使用一段包含人 hARGHase 基因序列的全部或部分的随机引物化 DNA 探针筛选。要完成对与人 hARGHase 序列至少有 70% 同源性的 DNA 插入序列的克隆的鉴定, 可以使用杂交温度为 55°C 的杂交液, 然后用 $0.5\times\text{SSC}$ 和 $0.1\%\text{SDS}$ 清洗。用这种方法识别的克隆的 DNA 插入序列可以进一步用 DNA 限制性内切酶分析和 DNA 测序来评价它与人

hARGHase 基因的相似性。组织表达的分布可以用上述的 Northern 印迹法技术分析。

人 hARGHase 同源物也可以用针对人 hARGHase 蛋白或多肽的抗体来识别。例如, 可以用标准的方法对商品化的或者用已知方法构建的, 来自细胞或者组织例如肝癌组织的表达文库进行筛选。将文库倒入平皿, 菌落转移到一张硝化纤维素膜上, 使表达的重组蛋白结合到膜上。然后就可以用特异性的人 hARGHase 抗体进行典型的抗体结合和检测。用这种方法所识别出克隆中的 DNA 插入序列, 可以进一步用 DNA 限制性内切酶分析和 DNA 测序进行分析以评价它与人 hARGHase 基因的相似性。新识别的基因的组织表达分布可以同样地按上述方法进行分析。

本发明的人 hARGHase 核苷酸全长序列或其片段通常可以用 PCR 扩增法、重组法或人工合成的方法获得。对于 PCR 扩增法, 可根据本发明所公开的有关核苷酸序列, 尤其是开放阅读框序列来设计引物, 并用市售的 cDNA 库或按本领域技术人员已知的常规方法所制备的 cDNA 库作为模板, 扩增而得有关序列。当序列较长时, 常常需要进行两次或多次 PCR 扩增, 然后再将各次扩增出的片段按正确次序拼接在一起。

一旦获得了有关的序列, 就可以用重组法来大批量地获得有关序列。这通常是将其克隆入载体, 再转入细胞, 然后通过常规方法从增殖后的宿主细胞中分离得到有关序列。

此外, 还可用人工化学合成的方法来合成有关序列。在本申请之前, 现有技术已完全可以通过先合成多个多核苷酸小片段, 然后再进行连接而获得编码本发明人 hARGHase 蛋白的核酸序列。然后, 可将该核酸序列引入本领域中各种现有的 DNA 分子(如载体)和细胞中。此外, 还可通过化学合成将突变引入本发明蛋白序列中。

除了用重组法产生之外, 本发明蛋白的片段还可用固相技术, 通过直接合成肽而加以生产(Stewart 等人, (1969) Solid-Phase Peptide Synthesis, WH Freeman Co., San Francisco; Merrifield J. (1963) J. Am Chem. Soc 85: 2149-2154)。在体外合成蛋白质可以用手工或自动进行。例如, 可以用 Applied Biosystems 的 431A 型肽合成仪(Foster City, CA)来自动合成肽。可以分别化学合成本发明蛋白的各片段, 然后用化学方法加以连接以产生全长的分子。

本发明蛋白的编码序列可用于基因定位。例如, 通过荧光原位杂交技术(FISH), 将 cDNA 克隆与分裂中期的染色体进行杂交, 可以准确地进行染色体定位。该技术可以使用短至约 500bp 的 cDNA; 也可以使用长至约 2000bp 或者更长的 cDNA。对于该技术, 可参见 Verma 等人, Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York(1988)。

一旦序列被定位于染色体上的某个精确位置, 将可以将序列在染色体上的物理

位置与遗传图谱数据相关联。这些遗传图谱数据是可以获得的，例如通过孟德尔(Mendelian)人遗传数据库(可通过 Johns Hopkins University Welch Medical Library 在网上获得)。然后，通过连锁分析来鉴定基因与已定位于同一染色体区域的疾病之间的相关性。

接着，有必要确定患病个体和健康个体之间的 cDNA 或基因组序列方面的差异。如果某一突变存在于部分或全部患病个体但不存在于正常个体，那么该突变可能就是该疾病的致病因素。

利用本发明的人 hARGHase 蛋白，通过各种常规筛选方法，可筛选出与人 hARGHase 发生相互作用的物质，或者受体、抑制剂或拮抗剂等。

本发明人 hARGHase 蛋白及其抗体、抑制剂、拮抗剂或受体等，当在治疗上进行施用(给药)时，可提供不同的效果。通常，可将这些物质配制于无毒的、惰性的和药学上可接受的水性载体介质中，其中 pH 通常为约 5—8，较佳地 pH 约为 6—8，尽管 pH 值可随被配制物质的性质以及待治疗的病症而有所变化。配制好的药物组合物可以通过常规途径进行给药，其中包括(但不限于)：肌内、腹膜内、皮下、皮内、或局部给药。

以本发明的人 hARGHase 蛋白为例，可以将其与合适的药学上可接受的载体联用。这类药物组合物含有治疗有效量的蛋白质和药学上可接受的载体或赋形剂。这类载体包括(但不限于)：盐水、缓冲液、葡萄糖、水、甘油、乙醇、及其组合。药物制剂应与给药方式相匹配。本发明的人 hARGHase 蛋白可以被制成针剂形式，例如用生理盐水或含有葡萄糖和其他辅剂的水溶液通过常规方法进行制备。诸如片剂和胶囊之类的药物组合物，可通过常规方法进行制备。药物组合物如针剂、溶液、片剂和胶囊宜在无菌条件下制造。活性成分的给药量是治疗有效量，例如每天约 1 微克/千克体重—约 5 毫克/千克体重。此外，本发明的多肽还可与其他治疗剂一起使用。

当本发明的人 hARGHase 蛋白多肽被用作药物时，可将治疗有效剂量的该多肽施用于哺乳动物，其中该治疗有效剂量通常至少约 10 微克/千克体重，而且在大多数情况下不超过约 8 毫克/千克体重，较佳地该剂量是约 10 微克/千克体重—约 1 毫克/千克体重。当然，具体剂量还应考虑给药途径、病人健康状况等因素，这些都是熟练医师技能范围内的。

表 2 为本发明的人 hARGHase 蛋白与 *Aquifex aeolicus* 的 ARGHase 蛋白的氨基酸序列(GenPept Accession No. AAC06773)的同源比较(FASTA)表。其中，相同的氨基

酸在两个序列之间用氨基酸单字符标出, 相似的氨基酸用“+”标出。

下面结合具体实施例, 进一步阐述本发明。应理解, 这些实施例仅用于说明本发明而不适用于限制本发明的范围。下列实施例中未注明具体条件的实验方法, 通常按照常规条件, 例如 Sambrook 等人, 分子克隆: 实验室手册 (New York: Cold Spring Harbor Laboratory Press, 1989) 中所述的条件, 或按照制造厂商所建议的条件。

实施例 1

人 hARGHase 基因的克隆

1. 组织分离 (肝癌组织 isolation)

肝癌组织来源于 5 个正常成年男性供体, 在死后四小时内取出肝癌组织, 立即置于液氮中冷冻保存。

2. mRNA 的分离 (mRNA isolation)

取出组织, 用研钵研碎, 加入盛有裂解液的 50ml 管, 充分振荡后, 再移入玻璃匀浆器内。匀浆后移至 50ml 新管, 抽提总 RNA (TRIzol Reagents, Gibco, NY, USA)。用甲醛变性胶电泳鉴定总 RNA 质量。用带 Oligo d(T) 的纤维素柱分离总 RNA 中的 mRNA, 定量。

3. cDNA 文库的构建 (Construction of cDNA library)

以 mRNA 为模板, 合成双链 cDNA, 反转录引物见 SEQ ID NO. 1。补平末端后, 加含 EcoRI 切点的接头, 接头序列分别见 SEQ ID NO. 2 和 3。磷酸化 EcoRI 末端后, 用 XhoI 限制性内切酶消化 1.5 小时, 再进行片段分离。过柱筛选长度 >500bp 的片段, 用酚-氯仿抽提, 乙醇沉淀, 无菌水溶解, 连接至 Uni-ZAP XR 载体 (Stratagene, CA9203, USA), 以 Zap-cDNA Gigapack III Gold Cloning Kit (Stratagene, CA9203, USA) 进行包装, 宿主菌使用 XL 1-Blue MRF⁻ (Stratagene, CA9203, USA) 细菌。涂板并测定滴度。

4. 测序及数据库建立 (Sequencing and Database Constructing)

挑选文库中有外源片段插入的克隆, 扩增后抽提质粒 (Qiagen, Germany), 用 T3 和 T7 作为 3' 端和 5' 端的通用引物, 采用终止物荧光标记 (Big-Dye, Perkin-Elmer, USA) 的方法, 在 ABI 377 测序仪 (Perkin-Elmer, USA) 上进行 EST 大规模测序。测序结果用 FACTURA 软件去除载体序列, 传输到 SUN Ultra 450 Server 上进行下一步的处理。所有的序列信息再用 GCG 软件包 (Wisconsin group, USA) 中的 BLAST 和 FASTA 软件搜索已有的数据库 (Genbank+EMBL), 将无同源性或同源性低于 95% 的序列视为

新基因建立数据库。

5. 基因的全长克隆(Cloning of Full-length cDNA)

在得到的新基因片段序列信息基础上, 进行 cDNA 全长克隆, 分两阶段进行:

(1) “电子克隆”(Electronic Cloning)

以新基因片段序列作为探针搜寻 dbEST 数据库, 将重叠序列>50bp, 同源性在 98% 以上的表达序列标签(Expressed Sequence Tag, 简称“EST”)序列认为同一序列(consensus sequence), 取出并用 AUTOASSEMBLER 软件进行拼接, 部分 EST 可以延伸探针序列。再用 STRIDER 软件分析被延伸的序列是否具有完整的开放阅读框架(Open Reading Frame, ORF), 用 BLAST 搜寻 Genbank 或 GenPept 以确定该序列在核苷酸和氨基酸水平上是否与其他物种有同源性, 以帮助判别所得到的基因全长完整性如何。通过电子克隆的方法, 通常可获取人 hARGHase 基因的全长序列。

(2) cDNA 末端快速扩增(Rapid Amplification of cDNA Ends, RACE)

如果通过“电子克隆”方法仍未得到完整的 cDNA 全长, 则在已有序列的 5' 或 3' 端设计引物, 在人类肝癌组织 Marathon-Ready cDNA 文库(Clontech Lab, Inc, USA)中进行长距离 PCR 反应。然后对 PCR 产物克隆、测序。用 AUTOASSEMBLER 及 STRIDER 软件分析被延长的序列有无完整的 ORF, 如无, 重复上述过程直至获得全长。

(3) RT-PCR

对于 5' 和 3' 端已知的序列, 如果中间尚有一段间隙(gap)无法从已有的公共数据库或自身数据库获得, 可考虑采用 RT-PCR 的方法。在序列 5' 端设计引物, 3' 端引物采用 Oligo-dT, 在肝癌组织总 RNA 库中进行扩增。然后对产物进行克隆、测序。最后拼接并获得全长。

通过组合使用上述 3 种方法, 获得了候选的人 hARGHase 蛋白的全长编码序列。在拼接得到全长(至少包含完整的开放读框)的基础上, 进一步设计引物 R1: 5'-TAGGGAACTGTGAGGGCG-3' (SEQ ID NO. 4) 为正向引物, 寡核苷酸 R2: 5'-CTGTCCCTTCATCCACACAAC-3' (SEQ ID NO. 5) 为反向引物, 以肝癌组织的总 RNA 为模板, 进行 RT-PCR 扩增, R1/R2 的 PCR 条件为 94°C 5 分钟, 随之以 94°C 30 秒、57°C 30 秒和 72°C 1 分钟进行 35 个循环, 最后以 72°C 延伸 5 分钟。电泳检测 PCR 扩增产物, 获得扩增片段长度为 1582 bp。然后按常规方法以 PCR 扩增产物进行克隆、测序, 获得 SEQ ID NO. 6 所示的序列。

实施例 2

人 hARGHase 基因的序列信息与同源性分析:
本发明新的人 hARGHase 全长 cDNA (GenBank Accession No. AF212236。在本申请之

前未公开过)的长度为 1582 bp, 详细序列见 SEQ ID NO. 6, 其中开放读框位于 51-1142 位核苷酸。根据全长 cDNA 推导出人 hARGHase 的氨基酸序列, 共 363 个氨基酸残基, 分子量 40071.57, pI 为 6.52。详细序列见 SEQ ID NO. 7。

将人 hARGHase 的全长蛋白质用 BLAST 程序在 Non-redundant GenBank CDS translations +PDB+ GenPept+Superdate+PIR 数据库中进行蛋白质同源性检索, 结果发现它与 *Aquifex aeolicus* 的 ARGHase 基因存在一定的同源性。在氨基酸水平上, 它与 *Aquifex aeolicus* 的 ARGHase 蛋白(GenPept Accession No. AAC06773)的第 18-325 位氨基酸残基有 30.0%的相同性和 59.6%的相似性(见表 2)。由上可见, 人 hARGHase 基因与 *Aquifex aeolicus* 的 ARGHase 基因在蛋白水平上存在较高的同源性, 属于同一家族并且两者在功能上也有很高相似性。

本发明的人 hARGHase 除了可作为该家族一员用于进一步的功能研究, 还可用于与其他蛋白一起产生融合蛋白, 比如与免疫球蛋白一起产生融合蛋白。此外, 本发明的人 hARGHase 还可以与该家族的其他成员进行融合或交换片段, 以产生新的蛋白。例如将本发明的人 hARGHase 蛋白的 N 端与 *Aquifex aeolicus* 的 ARGHase 蛋白的 N 端进行交换, 以产生新的活性更高或具有新特性的蛋白。

针对本发明人 hARGHase 的抗体, 用于筛选该家族的其他成员, 或者用于亲和纯化相关蛋白(如该家族的其他成员)。

此外, 本发明人 hARGHase 核酸(编码序列或反义序列)可以被引入细胞, 以提高人 hARGHase 的表达水平或者抑制人 hARGHase 的过度表达。本发明的人 hARGHase 蛋白或其活性多肽片段可以施用于病人, 以治疗或减轻因人 hARGHase 缺失、无功能或异常而导致的有关病症。此外, 还可以用基于本发明的核酸序列或抗体进行有关的诊断或预后判断。

由于本发明的人 hARGHase 蛋白具有源自人的天然氨基酸序列, 因此, 与来源于其他物种(如 *Aquifex aeolicus*)的同族蛋白相比, 预计在施用于人时将具有更高的活性和/或更低的副作用(例如在人体内的免疫原性更低或没有)。

实施例 3

人 hARGHase 基因的组织表达谱

电子 Northern 表达谱。按 Ton C. 等人的方法(Ton C et al., Biochem Biophys Res Commun 1997 Dec 18; 241(2): 589-594; Hwang DM, et al., Circulation 1997 Dec 16; 96(12): 4146-4203), 将人 hARGHase cDNA 序列在 GCG 软件包中的 dbEST 数据库中做 BLAST 检索, 在得到的人类 EST 中, 概率值 $<10e-10$ 、相同性 $>95\%$ 的 EST 有上百个, 可视为该基因在组织中的转录表达本, 由此得出表达该基因的组织谱,

发现其在结肠癌、神经上皮细胞、甲状旁腺肿瘤、胎盘、婴儿心脏、T 细胞淋巴瘤、睾丸、胎脑、骨骼肌等器官或组织中有表达,表明它在人体某些重要组织或器官中发挥着作用。

实施例 4

人 hARGHase 多肽的制备和提纯

在该实施例中,将全长的人 hARGHase 编码序列或片段构建入商品化的蛋白质融合表达载体之中,以表达和提纯重组蛋白。

将人 hARGHase 多肽以 GST 融合蛋白的形式在大肠杆菌中进行原核表达。

原核表达载体的构建,以及转化大肠杆菌

根据人 hARGHase 的全长编码序列(SEQ ID NO. 6),设计扩增出完整编码阅读框的引物(分别对应于编码序列 5' 和 3' 端的约 20 个以上核苷酸),并在正反引物上分别引入限制性内切酶位点(这根据选用的 pGEX-2T 载体而定),以便构建表达载体。以实施例 1 中获得的扩增产物为模板,经 PCR 扩增后,将人 hARGHase 基因在保证阅读框正确的前提下克隆至 pGEX-2T 载体(Pharmacia, Piscataway, NJ)。鉴定好的表达载体利用 CaCl_2 方法转入大肠杆菌 DH5 α , 筛选鉴定得到含有 pGEX-2T-hARGHase 表达载体的工程菌 DH5 α -pGEX-2T-hARGHase。

表达 GST-hARGHase 重组蛋白的工程菌的分离鉴定

挑取单菌落的 DH5 α -pGEX-2T-hARGHase 工程菌于 3ml 含 100 $\mu\text{g}/\text{ml}$ 氨苄青霉素的 LB 培养基中振荡培养过夜,按 1:100 的浓度吸取培养液于新的 LB 培养基(含 100 $\mu\text{g}/\text{ml}$ 氨苄青霉素)中培养约 3 小时,至 OD_{600} 达 0.5 后,加入 IPTG 至终浓度 1 mmol/L 继续于 37 $^{\circ}\text{C}$ 分别培养 0, 1, 2, 3 小时。取培养时间不同的 1ml 菌液离心,在细菌沉淀物中加入裂解液(2 \times SDS 上样缓冲液 50 μl , 蒸馏水 45 μl , 二巯基乙醇 5 μl),混悬细菌沉淀,沸水浴中煮 5 分钟,10000rpm 离心 1 分钟,上清加入 12%SDS-PAGE 胶中电泳。染色后观察预期分子量大小的蛋白量随 IPTG 诱导时间增加而增加的菌株即为表达 GST-hARGHase 融合蛋白的工程菌。

GST-hARGHase 融合蛋白的提取纯化

按上述方法诱导表达 GST-hARGHase 融合表达蛋白的工程菌 DH5 α -pGEX-2T-hARGHase。诱导后的细菌离心沉淀,按每 400ml 菌加入 20ml PBS 重悬细菌,超声破碎细菌。破菌完全的超声液按每毫升加入 20 微升的量加入 PBS 饱和的 50%谷胱甘肽 Sepharose 4B, 37 $^{\circ}\text{C}$ 振荡结合 30 分钟,10000 rpm 离心 10 分钟沉淀结合了 GST-hARGHase 的谷胱甘肽 Sepharose 4B, 弃上清。按每毫升超声液所得沉淀加入 100 μl

PBS 的量清洗两次,而后按每毫升超声液所得沉淀加入 10 μ l 还原型谷胱甘肽洗脱液,室温置 10 分钟,10000 rpm 离心 10 分钟,上清即为洗脱的融合蛋白。重复洗脱两次。洗脱的上清保存于-80℃,并进行 SDS-PAGE 电泳,检测纯化效果。在 40kDa 处的蛋白质条带即为人 hARGHase 蛋白。

实施例 5

人 hARGHase 蛋白或多肽在昆虫细胞中进行真核细胞表达

1. 人 hARGHase 杆状病毒表达载体的构建及转染 Sf9 昆虫细胞株

根据人 hARGHase 的全长编码序列(SEQ ID NO. 6),设计扩增出完整编码阅读框的引物,并在正反引物上分别引入限制性内切酶位点(这可视选用的载体而定),以便构建表达载体。以实施例 1 中获得的扩增产物为模板,经 PCR 扩增后,将人 hARGHase cDNA 在保证阅读框架的前提下克隆至 pVL1392 载体(Invitrogen, Carlsbad, CA)。鉴定好的表达载体 3 μ g,野生型线性杆状病毒 DNA(BaculoGold™ ACMNPV DNA, Pharmingen, San Diego, CA)1 μ g 和 Lipofection(Gibco-BRL, NY)25 μ l,加入 1ml 无血清的昆虫培养基中,振荡 15 秒混匀,室温孵育 15 分钟备用。取 1 ml (2×10^6) Sf9 昆虫细胞悬液于 60mm 组织培养板中,贴壁 1 小时后换转染培养基,室温孵育 15 分钟后弃培养基,加入前面制备好的 DNA 载体转染混合物,Parafilm 密封培养板,于室温 27℃ 振荡培养 4 小时,而后换完全培养基培养 3 天,收集上清备用。

2. 转入重组表达载体的昆虫细胞株的筛选鉴定

转染 3 天后的昆虫细胞用新鲜培养基制成细胞悬液(2×10^6 /1 ml),取 1 ml 细胞悬液置于 60mm 组织培养皿中,加入 3ml 培养基,100 μ l 收集的培养上清,贴壁 1 小时,弃 2ml 培养基,继续室温培养 1 小时,弃去所有的培养基,加入预热的含 20 μ l 4% X-gal 的 3ml 半固体培养基,培养 5-7 天后挑取白色细胞克隆于 96 孔培养板中培养 3-5 天,而后吸取上清感染 Sf9 昆虫细胞。

收集感染的细胞进行 Western 鉴定。将细胞裂解后进行 SDS-PAGE 电泳,电泳后的胶于 Pharmacia 的 Multiphor II 半干电转移仪中将蛋白质转印到硝酸纤维膜上,将硝酸纤维膜置于封闭液中封闭 1 小时,而后于抗人 hARGHase 的抗体溶液中封闭 1 小时,TBS 液振荡清洗 5 分钟共 2 次,而后将膜置于生物素标记的抗人 hARGHase 一抗的第二抗体溶液中振荡 1 小时,TBS 清洗,加入亲和素-碱性磷酸酶复合物反应 30 分钟,TBS 清洗 2 次,加入新鲜配制的显色液显色观察蛋白条带。

挑取高表达人 hARGHase 的 Sf9 细胞克隆。

3. 人 hARGHase 蛋白的提取纯化

用高表达人 hARGHase 的 Sf9 细胞克隆的上清大量感染 Sf9 细胞, 感染 48 小时后收集细胞, PBS 洗涤。每 2×10^8 细胞加入 20 ml 细胞裂解液 (0.5% Triton X-100, 20 mM Na_3PO_4 (磷酸钠, pH7.8), 500mM NaCl, 1mM Na_3VO_4 (钒酸钠), 1mM Pefabloc, 1 $\mu\text{g}/\text{ml}$ 胃蛋白酶抑制剂, 亮抑蛋白酶肽和抑蛋白酶肽) 破细胞, $12000 \times g$ 离心 20 min 去除细胞碎片, 上清按每 2×10^8 的细胞加入 2 ml NTA-琼脂糖 (Qiagen, Germany), 4°C 吸附 1 小时。而后用含 100 nM 咪唑的 His 缓冲液洗涤两次, 用含 20mM N,N'-二哌嗪, 500 mM NaCl, 300mM 咪唑的缓冲液洗脱以得到纯化的蛋白。洗脱液保存于 4°C , 并进行 SDS-PAGE 电泳检测提取的人 hARGHase 蛋白的纯度。在 40kDa 处的蛋白质条带即为人 hARGHase 蛋白。

实施例 6

抗人 hARGHase 抗体的制备

1. 免疫小鼠的 ARGHase 和脾细胞的制备: 将实施例 5 和实施例 6 中获得的人 hARGHase 蛋白用层析法进行分离后备用, 也可以用 SDS-PAGE 凝胶电泳法进行分离, 将电泳条带从凝胶中割下, 并用等体积的完全 Freund's 佐剂乳化。取 6-8 周龄 Balb/C 雌鼠的 ARGHase, 用 50-100 $\mu\text{g}/0.2\text{ml}$ 乳化过的蛋白, 对小鼠的 ARGHase 进行腹膜内注射。14 天后, 用非完全 Freund's 佐剂乳化的同样抗原对小鼠的 ARGHase 以 50-100 $\mu\text{g}/0.2\text{ml}$ 的剂量再加强免疫一次, 3-5 天后用于融合。其中, 脾细胞制备见鄂征主编, 《组织培养和分子细胞学技术》, 北京出版社, 第 210 页。

2. 按《组织培养和分子细胞学技术》(同上), 第 371 页中的方法, 制备饲养细胞。

3. 按《组织培养和分子细胞学技术》(同上), 第 213 页中的方法, 进行细胞融合。

4. 抗体的检测: 在细胞融合 10-15 天后, 需逐孔进行检查, 一旦发现旺盛的杂交细胞集落生长, 就应用人 hARGHase 蛋白做抗体活性的初步筛选, 常用的方法有: 免疫荧光试验、发射免疫试验 (RIA)、酶联免疫吸附试验 (ELISA)。检查出抗体活性的孔后, 立刻进行克隆培养, 并分离出抗体。

本发明涉及的序列及记号分列如下:

(1) SEQ ID NO. 1 的信息

(i) 序列特征:

- (A) 长度: 50bp
- (B) 类型: 核苷酸
- (C) 链性: 单链
- (D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

(ix) 序列描述: SEQ ID NO. 1

GAGAGAGAGAGAGAGAGAGAACTAGTCTCGAGTTTTTTTTTTTTTTTTTTT 50

(2) SEQ ID NO. 2 的信息

(i) 序列特征:

- (A) 长度: 13bp
- (B) 类型: 核苷酸
- (C) 链性: 单链
- (D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

(ix) 序列描述: SEQ ID NO. 2

AATTCGGCAC GA G

13

(3) SEQ ID NO. 3 的信息

(i) 序列特征:

- (A) 长度: 9bp
- (B) 类型: 核苷酸
- (C) 链性: 单链
- (D) 拓扑结构: 线性

(ii) 分子类型: 寡核苷酸

(ix) 序列描述: SEQ ID NO. 3

GCCGTGCTC

9

(4) SEQ ID NO. 4 的信息

(i). 序列特征:

- (A) 长度: 19bp
- (B) 类型: 核苷酸

(C) 链性: 单链

(D) 拓扑结构: 线性

(ii). 分子类型: 寡核苷酸

(ix). 序列描述: SEQ ID NO. 4

TAGGGAACTGTGAGGGCG

19

(5) SEQ ID NO. 5 的信息

(i). 序列特征:

(A) 长度: 21bp

(B) 类型: 核苷酸

(C) 链性: 单链

(D) 拓扑结构: 线性

(ii). 分子类型: 寡核苷酸

(ix). 序列描述: SEQ ID NO. 5

CTGTCCCTTCATCCACACAAC

21

(6) SEQ ID NO. 6 的信息

(i) 序列特征:

(A) 长度: 1582bp

(B) 类型: 核苷酸

(C) 链性: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 核苷酸

(ix) 序列描述: SEQ ID NO. 6

```
1 TAGGGAACT GTGAGGGCGG CACCGGAAGT GGCGAGCAGT CTGCGCGCGG
51 ATGGCCCAGC GGCGATGGCG GCACGGCAGG TTGGAGGGGC TGGCGCGGCC
101 GCTCCCTCTC GCGCTCCGA GGCTGCCTGG CTGGCGCGCT GCTCGGGGAC
151 TGCGTGGGCT CCTTCTACGA GGCCACGAC ACCGTCGACC TGACGTCATC
201 GTGCGTCATG TCCAGAGTCT GGAGCCGGAC CCCGTA CTGC CCCGGAAGT
251 GAGCGGACAG AAGCCTTTGT ACTACACAGA TGACACAGCC ATGGCCAGGG
301 CCCTGGTGCA GTCCCTGCTA GCCAAGGAGG CCTTTGACGA GGTGGACATG
351 GCTCACAGAT TTGCTCAGGA GTACAAGAAA GACCCTGACA GGGGCTATGG
```

G041300

```

401  TGCTGGAGTA GTCAGTGTCT TCAAGAAGCT CCTGAACCCC AAATGTCGCG
451  ATGTCTTTGA GCCTGCCCGG GCCCAGTTTA ACGGGAAAGG CTCCTATGGC
501  AATGGAGGTG CCATGCGGGT GGCTGGCATC TCCCTGGCCT ATAGCAGTGT
551  CCAGGATGTG CAGAAGTTTG CCCGGCTCTC GGCCCAGCTG ACACACGCCT
601  CCTCCCTGGG TTACAATGGC GCCATCCTGC AGGCCCTGGC TGTGCACCTG
651  GCCTTGCAGG GCGAGTCTTC CAGCGAGCAC TTTCTCAAGC AACTCCTGGG
701  CCACATGGAG GATCTGGAGG GTGATGCCCA GTCCGTCTTG GATGCCAGGG
751  AGTTGGGCAT GGAGGAGCGT CCATACTCCA GCCGCCTGAA GAAGATTGGA
801  GAGCTTCTAG ACCAGGCATC GGTGACCAGG GAGGAAGTGG TGTCTGAGCT
851  AGGGAATGGC ATTGCTGCCT TTGAGTCGGT ACCCACCGCC ATCTACTGCT
901  TCCTACGCTG CATGGAGCCA GACCCTGAGA TCCCTTCTGC CTTCAATAGC
951  CTCCAAAGGA CTCTCATTTA TTCCATCTCA CTTGGTGGGG ACACAGACAC
1001 CATTGCCACC ATGGCTGGGG CCATTGCTGG TGCCTACTAT GGGATGGATC
1051 AGGTGCCAGA GAGCTGGCAG CAAAGCTGTG AAGGCTACGA GGAGACAGAC
1101 ATCCTGGCCC AAAGCCTGCA CCGTGTCTTC CAGAAGAGTT GATGAGGGCT
1151 ACAGCTGTTG GGGCTCTGCA GGTCCCCTGG GACACACTAC AGCTCCAATC
1201 AGAAACCCTG CGCTTCCTTG AGTGTGGCTT CCCACTTTTC CTGCATTGTG
1251 GAGCTGACTG AGTACACCGG TGAGGCTGGG GTCTCTGCAG GGGAGGTCAC
1301 TGGAACAGCG AGCAAGGGAC TGGTGCCTCG CTGGTGCTGG GTCTCTGGTT
1351 TGCTGCAGAG CCGTAGGACA CTCCTGGCTC CTCAGTAGGA CAGACAGACG
1401 CAGGCGGGTT TATTTTGGAG GGTACTTGT GGCATTTTCC TGTATTGTCT
1451 TGGACATGGG ATGTGGGGAG GTGGAAATGA TGAGCAGTAG CATCATTTCT
1501 CCCTGTTGGG TTTTAGCCAG TTTGCCAGCA AGCGCATCCT AGCAGGGTCC
1551 CCGAGCAGCA GGTGTGTGG ATGAAGGGAC AG

```

(7) SEQ ID NO. 7 的信息

(i) 序列特征:

(A) 长度: 363 氨基酸

(B) 类型: 氨基酸

(C) 链性: 单链

(D) 拓扑结构: 线性

(ii) 分子类型: 多肽

(ix) 序列描述: SEQ ID NO. 7:

1 MAQRRWRHGR LEGLARPLPL ALPRLPGWRA ARGLRGLLLR GPRHRRPDVI
51 VRHVQSLEPD PVLPREVSGQ KPLYTTDDTA MARALVQSLL AKEAFDEVDM
101 AHRFAQEYKK DPDRGYGAGV VTVFKLLNP KCRDVFEPAR AQFNGKGSYG
151 NGGAMRVAGI SLAYSSVQDV QKFARLSAQL THASSLGYNQ AILQALAVHL
201 ALQGESSSEH FLKQLLGME DLEGDAQSVL DARELGMEER PYSSRLKKIG
251 ELLDQASVTR EEVVSSELGNG IAAFESVPTA IYCFRLCMEP DPEIPSAFNS
301 LQRTLIYSIS LGGDTDTIAT MAGAIAGAYY GMDQVPESWQ QSCEGYEETD
351 ILAQSLHRVF QKS

表 2

30.0% identity in 307 aa overlap, 59.6% similarity in 307 aa overlap

h. pep	30	40	50	60	70	80
	RAARGLRGLLLRGPRHRRPDVIVRHVQSLEPDVLPREVSGQKPLYTTDDTAMARALVQS					
a. pep	VGDALGKSVEDITEEEVFIFYGDRIRDFVTPHPSSP--AYGQLPEETSDETTIFRLLES					
	20	30	40	50	60	70
h. pep	90	100	110	120	130	140
	LLAKEAFDEVDMAHRFAQEYKKDPDRGYGAGVVTVFKLLNPKCRDVFEPARAQFNGKGS					
a. pep	LVEKKALDVRDFLNLMEWYDREEMHRYPDPMLLTAIDLSS---RGI-NPST---HGLTS					
	80	90	100	110	120	
h. pep	150	160	170	180	190	200
	YNGGAMRVAGISLA-YSSVQDVQKFARLSAQLTHASS-LGYNGAILQALAVHLALQGES					
a. pep	FSVEGILRSVALGLYHYYNPELAAEGSKVVSLLTHRSVDISDASAILGALIAHL-VRGDF					
	130	140	150	160	170	180
h. pep	210	220	230	240	250	260
	SSEHFLKQLLGHEMEDLEGDAQSVLDARELGMEERPYSSRLKKIGELLDAQSVTREEVVSE					
a. pep	YLEDF-REKLRLIETLKDYAK-----EEK-HKKILDRVAELL-MESADLETAINT					
	190	200		210	220	230
h. pep	270	280	290	300	310	320
	LGNGIAAFESVPTAIYCFRLCME-PDPEIPSAFNSLQRTLIYSISLGGDTDTIATMAGAI					
a. pep	LGNGTFAFEAFPLALYIFLSNVENPEEALFQAVNSYGE-----FGGDTDAIGFLVGAM					
	240	250	260	270	280	
h. pep	330	340	350	360		
	AGAYYGMDQVPESWQQSCEGYEETDILAQSLHRVFQKS					
a. pep	LGAYYGEEAIPYHLRENVENSQKLRELAERLYEVVEKQVGFSQ					

290

300

310

320

330

h. pep: 人 hARGHase 的蛋白序列

a. pep: *Aquifex aeolicus* ARGHase 蛋白序列 (GenPept Accession No. AAC06773)